

Data Owner 1:

Walker 31.07 xxx
Sophia 11.04 xxx
Clerk 11.12 xxx
Jana 21.11 xxx

Bloom Filter
based
encoding

10101 xxx
11001 xxx
01101 xxx
10010 xxx

Exchange
Encoding Secrets

Data Owner 2:

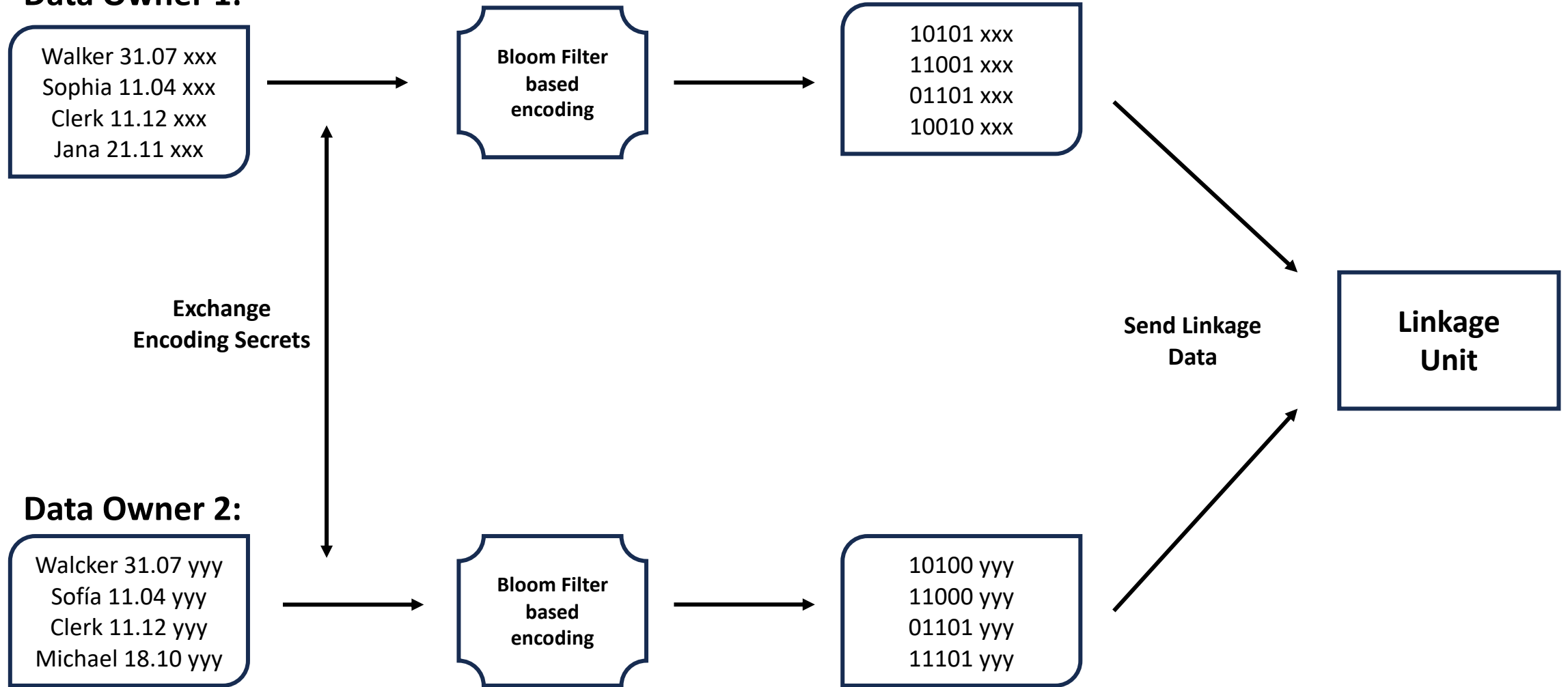
Walcker 31.07 yyy
Sofia 11.04 yyy
Clerk 11.12 yyy
Michael 18.10 yyy

Bloom Filter
based
encoding

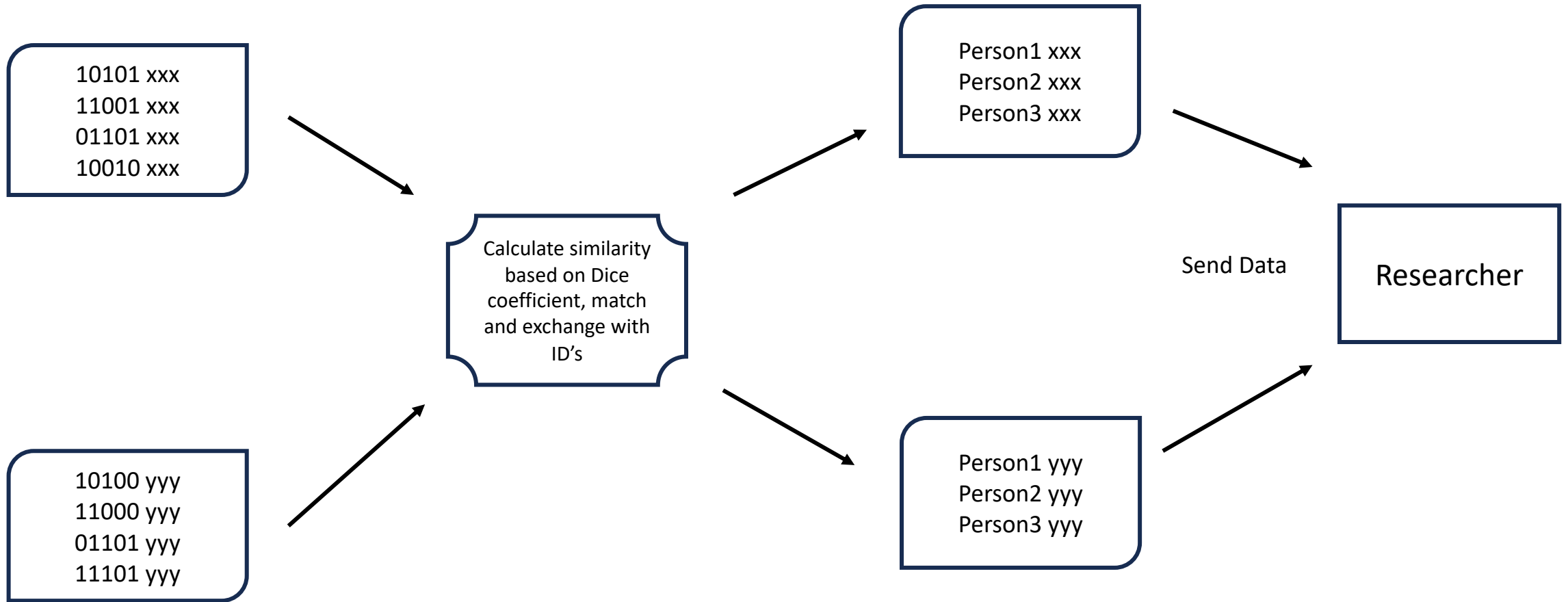
10100 yyy
11000 yyy
01101 yyy
11101 yyy

Send Linkage
Data

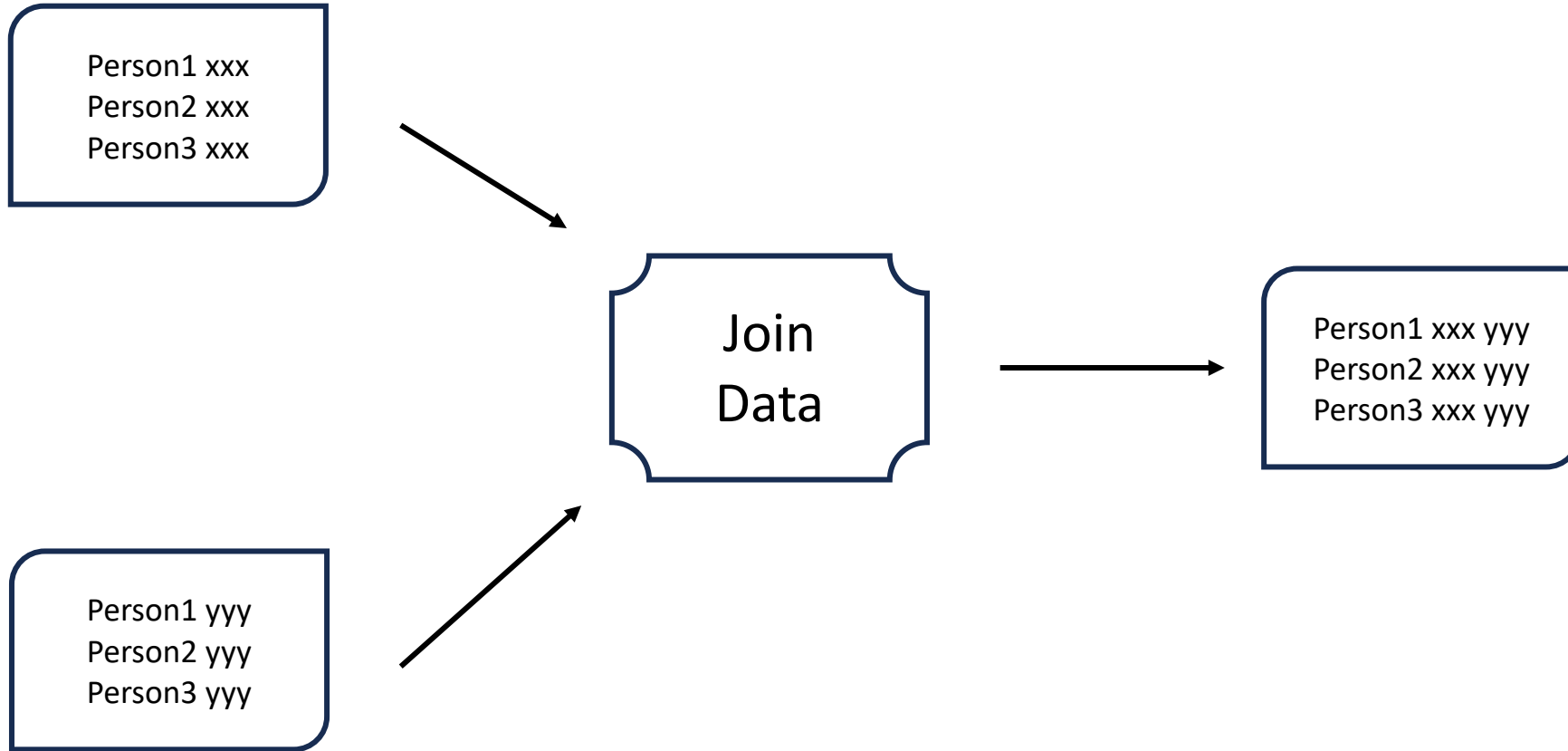
Linkage
Unit



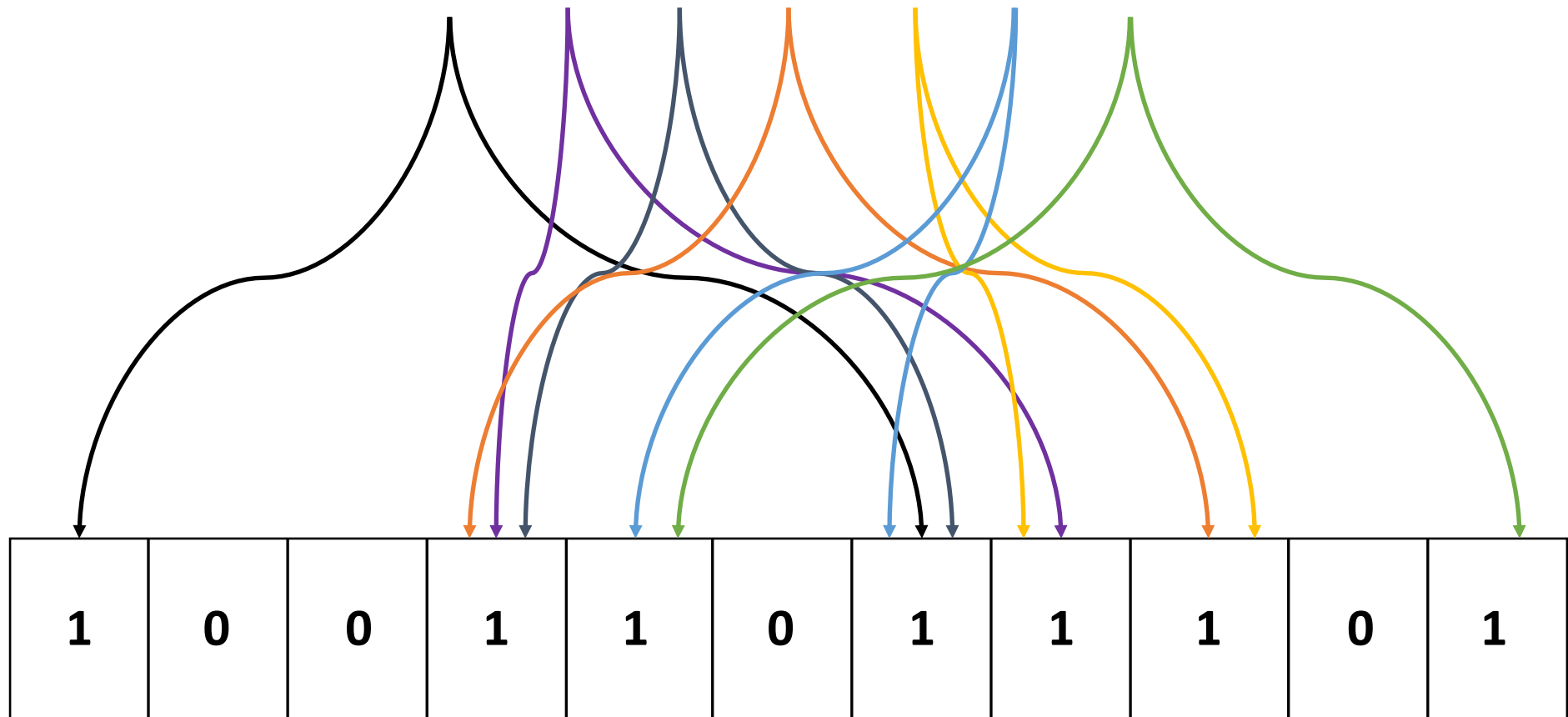
Linkage Unit:



Researcher:



{ {en}, {nc}, {co}, {od}, {di}, {in}, {ng} }



Index	Element	S	S'
1	1	1	1
2	2	0	1
3	3	1	1
4	4	1	0

π_1

Index	Element	S	S'
1	4	1	0
2	3	1	1
3	2	0	1
4	1	1	1

π_2

Index	Element	S	S'
1	3	1	1
2	4	1	0
3	2	0	1
4	1	1	1

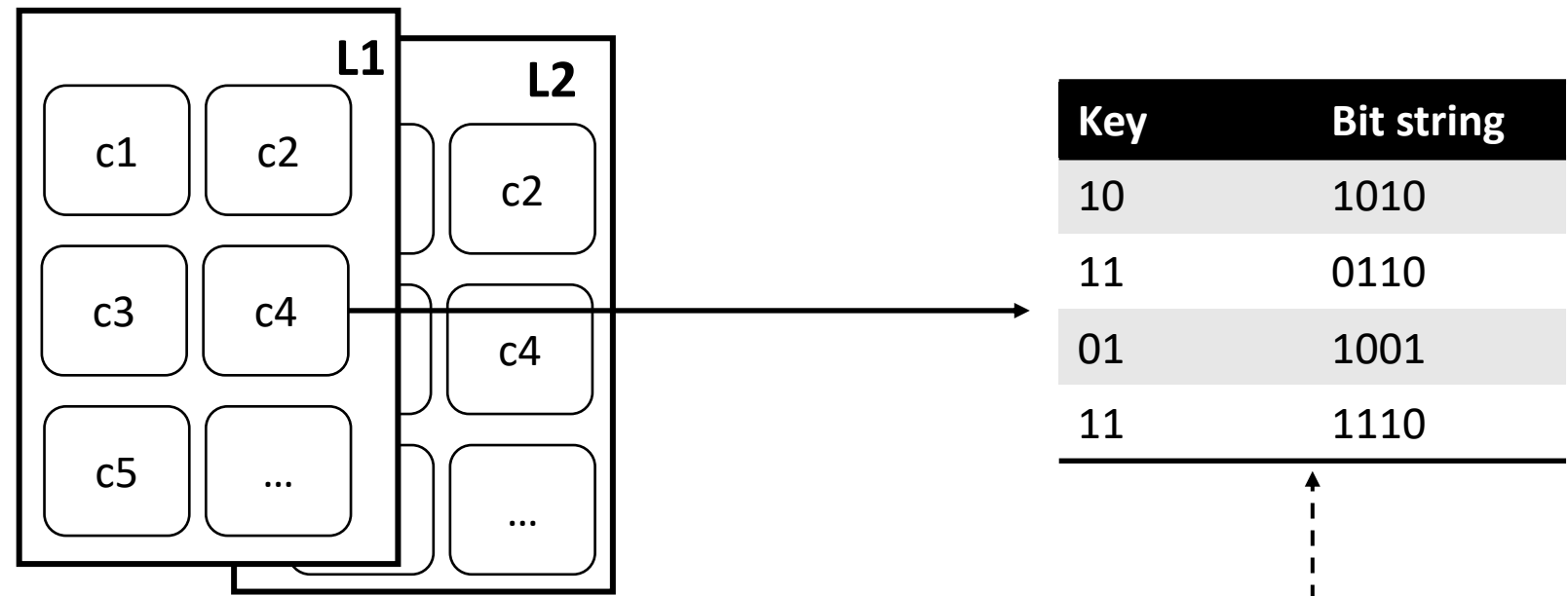
Estimated Jaccard Similarity

$$= [1(\pi_1(S_1)=\pi_1(S'_1)) + 1(\pi_2(S_1)=\pi_2(S'_1))] / 2$$

$$= (1 * 0 + 1 * 1) / 2$$

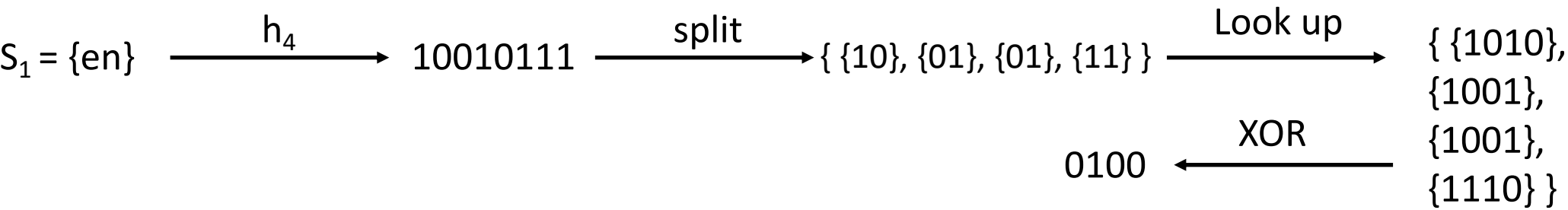
$$= 1/2$$

1) Initialization



$S = \{ \{en\}, \{nc\}, \{co\}, \{od\}, \{di\}, \{in\}, \{ng\} \}$

2) Hashing Process



$$\mathcal{Q}_1 = \{\text{pe, et, te, er}\}$$

\mathcal{H}_1	0	1	0	1	0	1	1	0
\mathcal{H}_2	1	1	0	0	1	0	1	0
\mathcal{H}_3	1	0	0	1	1	0	1	0
\mathcal{H}_4	0	0	1	1	0	1	0	1

$$\mathcal{G}(\mathcal{B}_p) \downarrow \downarrow \downarrow \downarrow \downarrow \downarrow \downarrow \downarrow \downarrow$$

$$E_1 = \{ 53, 113, 42, 7, 256, 87, 101, 21 \}$$

$$\mathcal{Q}_2 = \{\text{pe, et, te}\}$$

0	1	0	1	0	1	0	0
1	1	0	0	1	0	0	0
1	0	0	1	1	0	0	0
0	0	0	1	0	1	0	1

$$\downarrow \downarrow \downarrow \downarrow \downarrow \downarrow \downarrow \downarrow$$

$$E_2 = \{ 53, 113, 7, 256, 87, 21 \}$$

$$Sim_J(\mathcal{Q}_1, \mathcal{Q}_2) = \frac{|\mathcal{Q}_1 \cap \mathcal{Q}_2|}{|\mathcal{Q}_1 \cup \mathcal{Q}_2|}$$

$$= 0.75$$


$$Sim_J(E_1, E_2) = \frac{|E_1 \cap E_2|}{|E_1 \cup E_2|}$$

$$= 0.75$$

$S = \{ \{en\}, \{nc\}, \{co\}, \{od\}, \{di\}, \{in\}, \{ng\} \}$

hash

	l1	l2	l3	l4
k_1	1	1	1	0
k_2	1	0	0	0
k_3	0	0	1	0
k_4	1	1	0	0

$h'_1 \downarrow$ $h'_2 \downarrow$ $h'_3 \downarrow$ skip 

22 **13** **8**

result

$S = \{ \{22\}, \{13\}, \{8\} \}$

Data Owner 1:

Phillip 31.07 xxx
Sophia 11.04 xxx
Clerk 11.12 xxx
Jana 18.10 xxx

Bloom Filter
based
encoding

10101 xxx
11001 xxx
01101 xxx
10010 xxx

Data Owner 2:

Fillip 31.07 yyy
Sofía 11.04 yyy
Clerk 11.12 yyy
Michael 18.10 yyy

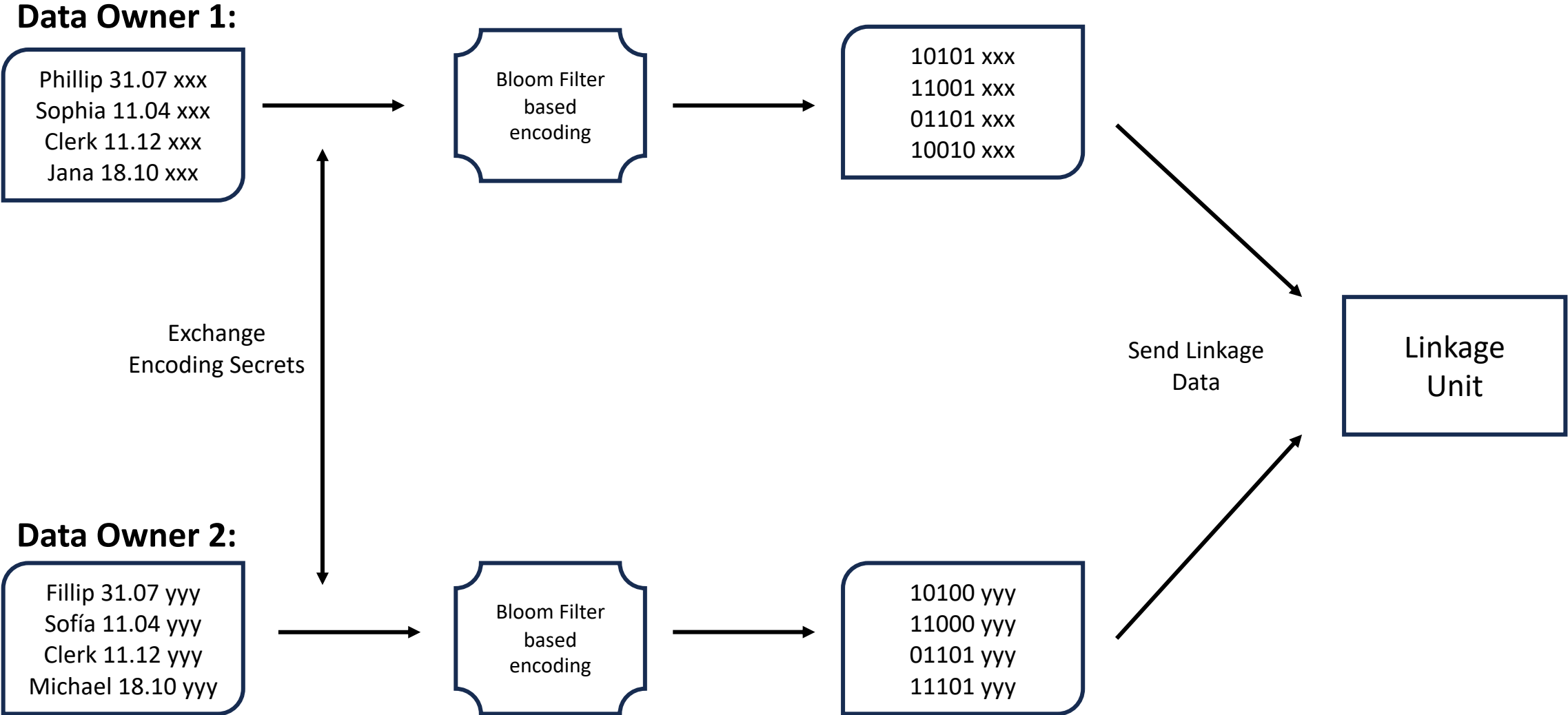
Bloom Filter
based
encoding

10100 yyy
11000 yyy
01101 yyy
11101 yyy

Linkage
Unit

Exchange
Encoding
Secrets

Send Linkage
Data

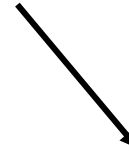
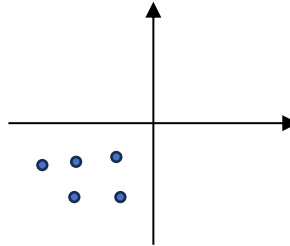


Linkage Unit:

10101 xxx
11001 xxx
01101 xxx
10010 xxx



Construct
similarity graph
and embeddings



Re-identify
records using
bipartite graph
matching



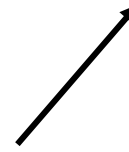
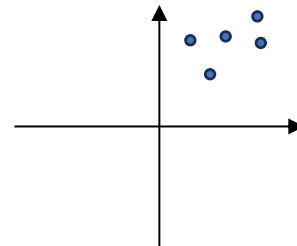
Sophia 11.04 xxx
Clerk 11.12 xxx
Jana 21.11 xxx

Public Phonebook:

Sophia 11.04
Clerk 11.12
Jana 21.11
Michael 18.10



Mimic BF
encoding and
construct
similarity graph
with embeddings



Re-identify
records using
bipartite graph
matching



Sophia 11.04 xxx
Clerk 11.12 xxx
Jana 21.11 xxx

Data Owner 1:

Walker 31.07 xxx
Sophia 11.04 xxx
Clerk 11.12 xxx
Jana 18.10 xxx

Bloom Filter
based
encoding

10101 xxx
11001 xxx
01101 xxx
10010 xxx

Data Owner 2:

Walcker 31.07 yyy
Sofia 11.04 yyy
Clerk 11.12 yyy
Michael 18.10 yyy

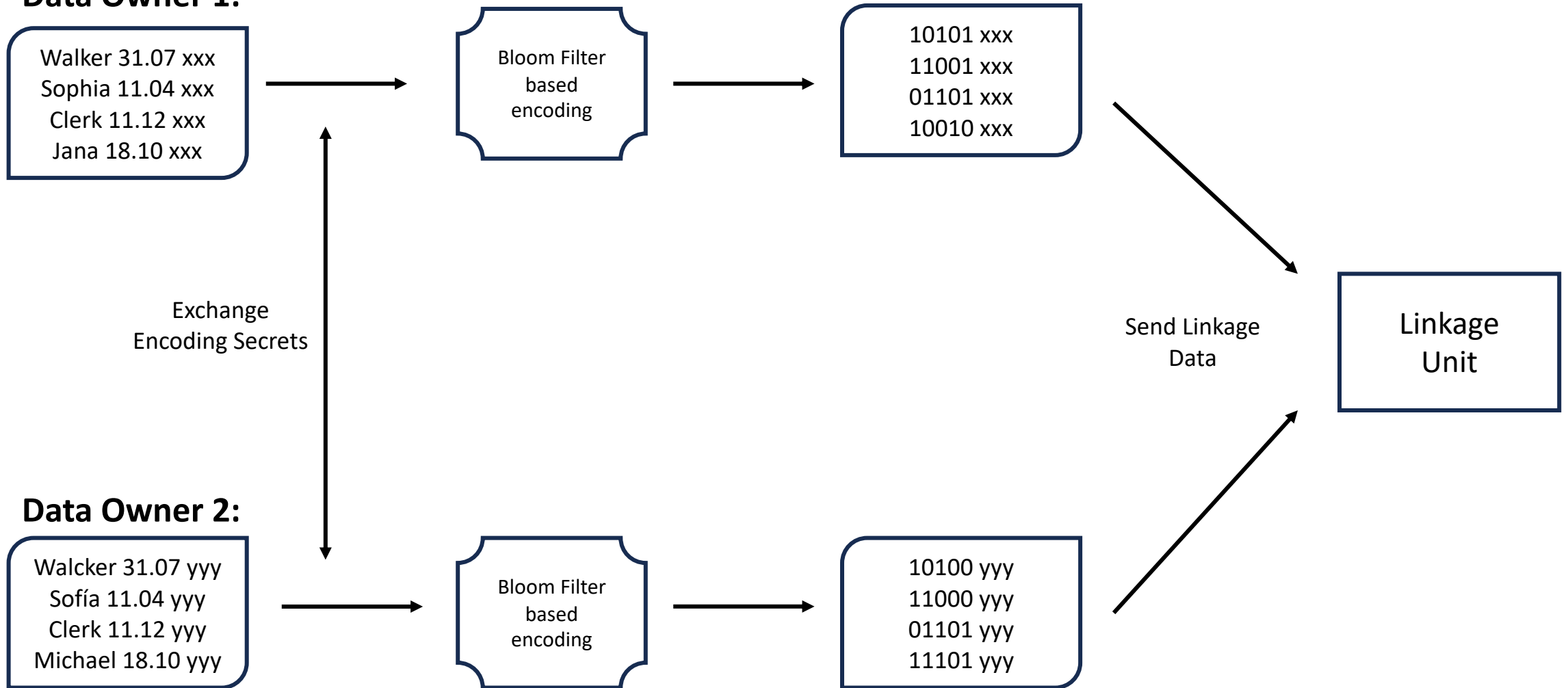
Bloom Filter
based
encoding

10100 yyy
11000 yyy
01101 yyy
11101 yyy

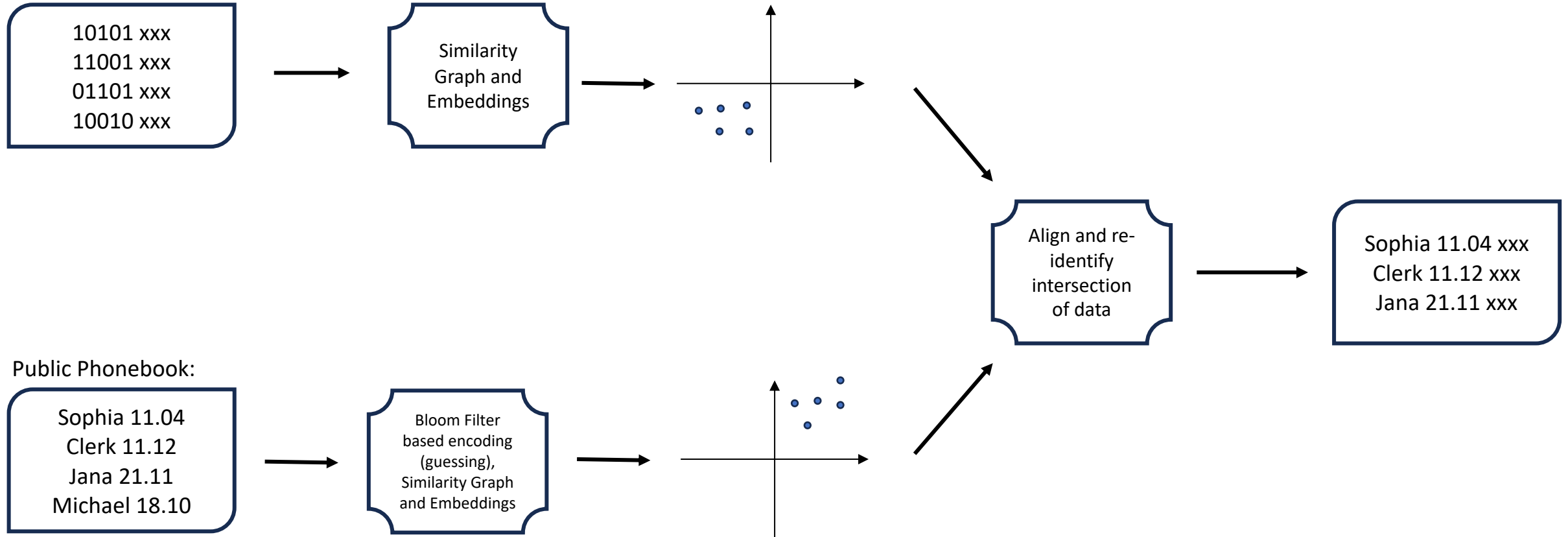
Linkage
Unit

Exchange
Encoding
Secrets

Send Linkage
Data



Linkage Unit:



Linkage Unit:

Known:

10101 Sophia 11.04 xxx

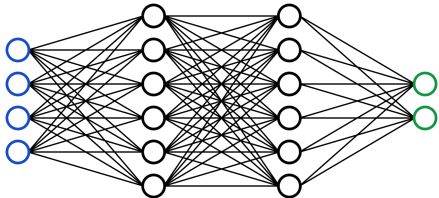
11001 Clerk 11.12 xxx

01101 Jana 21.11 xxx

Train Neural
Network

Unknown:

10010 xxx



10010 Walker 31.07 xxx

Graph Matching Attack



Data Preparation



Hyperparameter
Optimization



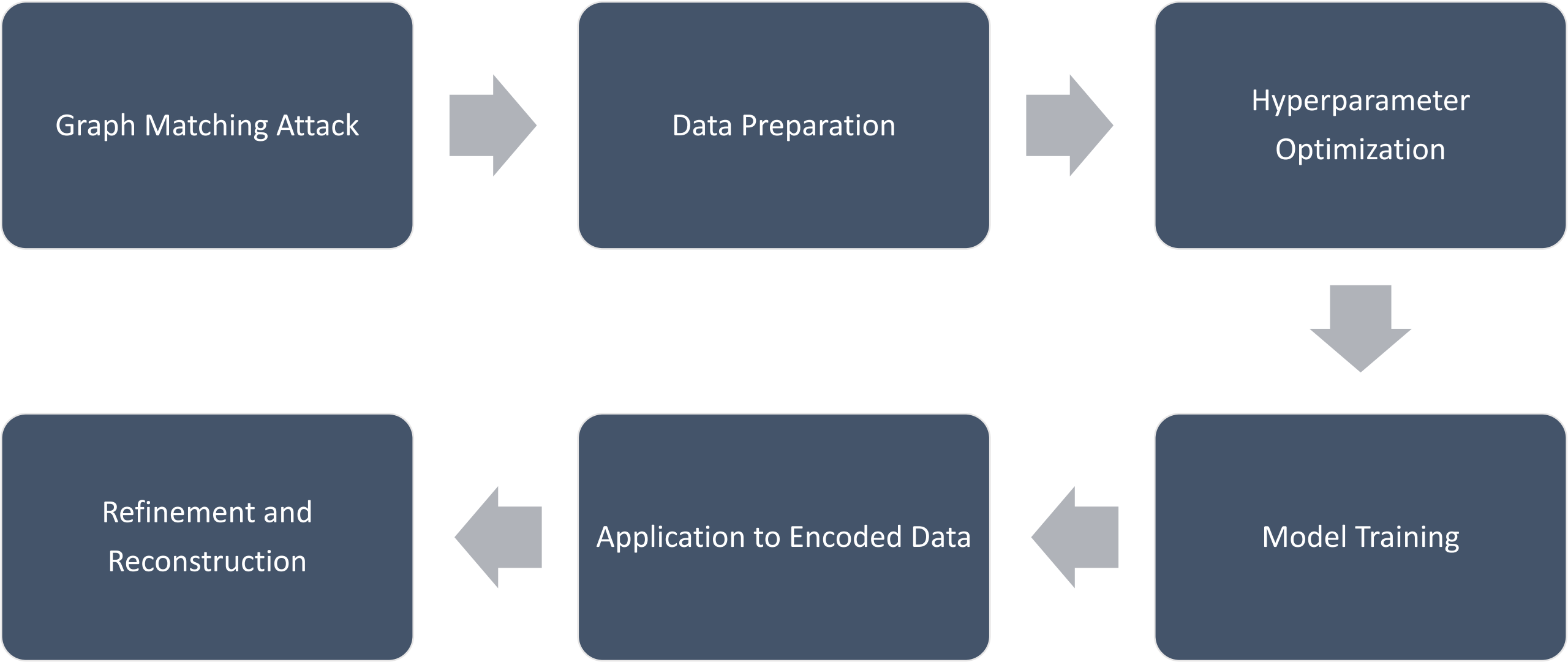
Refinement and
Reconstruction



Application to Encoded Data



Model Training



Input layer

Hidden layers

Output layer

i

h_1

h_2

h_n

o

Input 1

Input 2

⋮

Input n

⋮

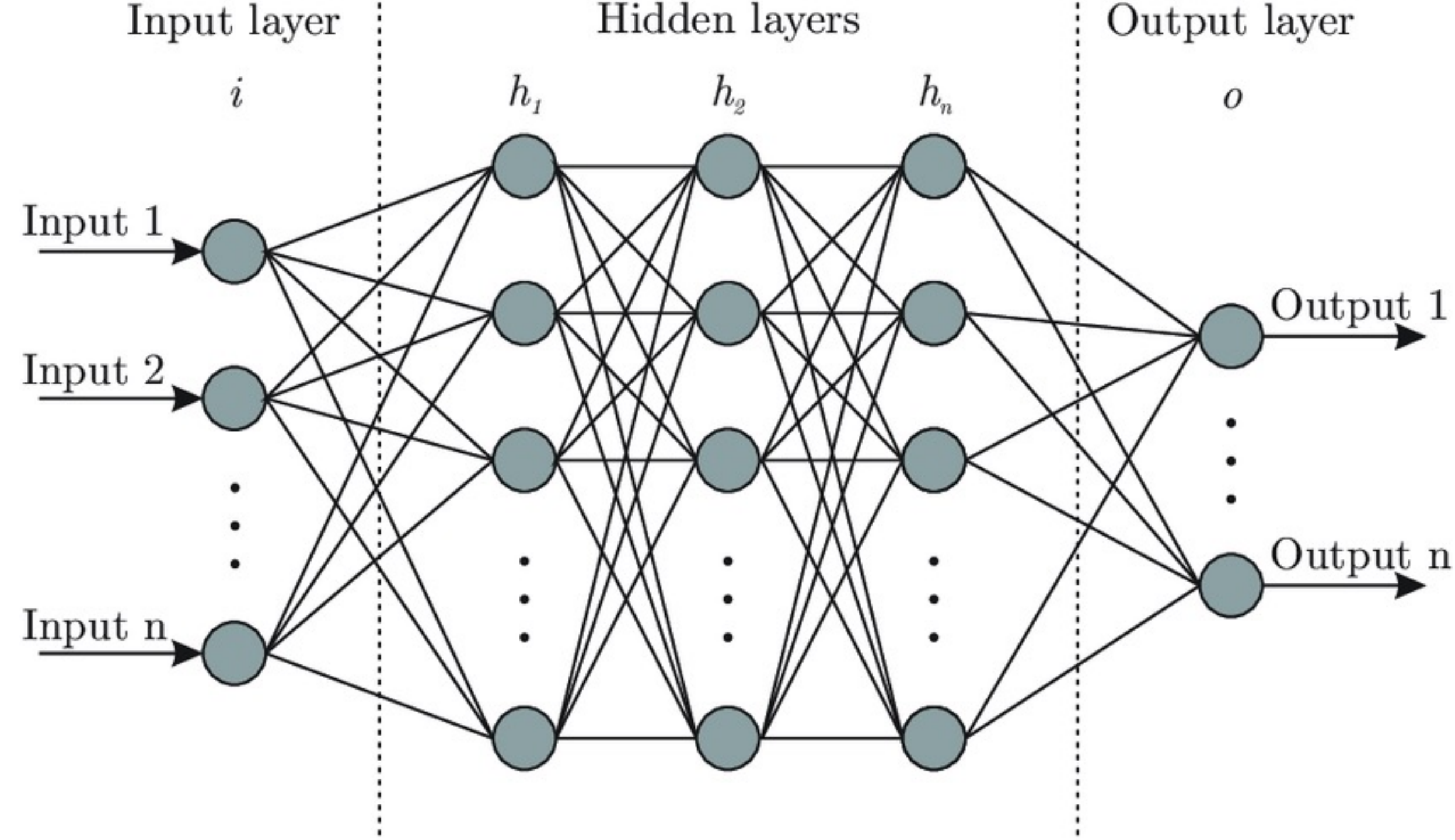
⋮

⋮

Output 1

⋮

Output n



Input

